

Guidelines for the Operation of the Census Bureau Research Data Centers

March 15, 2006

Note: This document combines and updates a document entitled “Guidelines for Providing Researchers Access to Title 13 Data at Research Data Centers” dated September 29, 1997, and a document entitled “Guidelines and Procedures for Hosting ‘Internal Projects’ at the Research Data Centers” dated March 12, 2004.

Background:

Since 1994, the Census Bureau has operated remote Research Data Centers (RDCs) where researchers with approved projects of benefit to the Census Bureau can access confidential Census Bureau (Title 13) and administrative data. The RDCs are operated by the Census Bureau’s Center for Economic Studies (CES) – the program manager – in cooperation with partner institutions (academic and non-profit organizations).

All RDCs must have a Census Bureau employee present and comply with strict federal government, Census Bureau, and Internal Revenue Service (IRS) information technology and physical security guidelines. All researchers accessing data at an RDC must be sworn to uphold the confidentiality of information provided by respondents. Title 13, U.S.C., Section 23(c), permits the Census Bureau to use temporary staff, **but only if such staff are assisting the Census Bureau in performing the Bureau’s work under Title 13 U.S.C.; and** only if such staff are sworn to observe the confidentiality provisions of Title 13, Section 9. Access to confidential data at a Census Bureau RDC can only be granted to persons with Special Sworn Status who have completed the necessary forms and swear to protect the data.

Guidelines for the operation of the RDCs were issued on September 29, 1997. After eight years of experience using those guidelines, the Census Bureau is issuing a revised set of guidelines. No significant changes are made, though the interpretation of various guidelines has been clarified.

Guidelines

I. Project Review and Approval

Research proposals submitted to CES for the use of confidential Census Bureau data are reviewed and judged against five major review standards that demonstrate:

- A benefit to Census Bureau programs conducted under Title 13;
- Scientific merit in that the research will contribute to existing knowledge;
- A clear need for non-public data;
- Feasibility of success; and
- Acceptance of all confidentiality protection and disclosure avoidance review requirements.

Each of the five standards is discussed in more detail below.

A. Benefit to Census Bureau programs.

Proposals must demonstrate that the research will provide benefits to data programs the Census Bureau conducts under Title 13 of the U.S. Code. Research proposals using only Title 13 data have to satisfy at least one of the 13 Census Bureau benefits documented in DS-002, *Articulating the Title 13 Benefits of Census Bureau Projects (DS-002)*, dated October 10, 2002, listed below. Census Bureau staff determine whether proposals using only Title 13 data provide adequate programmatic benefits.

Projects using, in addition, Title 26 Federal Tax Information (FTI) face a stricter standard. They must demonstrate that the project's *predominant* purpose is to benefit Census Bureau programs authorized under Chapter 5 of Title 13 and they must satisfy at least one of the nine benefits numbered 5 through 13 below. These nine benefits are those agreed to by the Census Bureau and the IRS in the *Criteria for the Review and Approval of Census Projects that Use Federal Tax Information*, dated September 15, 2000. Census Bureau staff and IRS staff jointly determine whether the predominant purpose of a project proposing to use FTI is to benefit Census Bureau programs conducted under Title 13.

Articulating the Title 13 Benefits of Census Bureau Projects

1. Evaluating concepts and practices underlying Census Bureau statistical data collection and dissemination practices, including consideration of continued relevance and appropriateness of past Census Bureau procedures to changing economic and social circumstances;
2. Analyzing demographic and social or economic processes that affect Census Bureau programs, and that evaluate improvements to the quality of products issued by the Census Bureau;
3. Evaluating or analyzing public programs, public policy, and/or demographic, economic, or social conditions to identify potential complementary datasets, improve data quality, enhance data collection techniques or develop innovative estimation procedures;
4. Conducting or facilitating census and survey data collection, processing or dissemination, including through activities such as administrative support, information technology support, program oversight, or auditing under appropriate legal authority;
5. Understanding and/or improving the quality of data produced through a Title 13, Chapter 5 survey, census, or estimate;
6. Leading to new or improved methodology to collect, measure, or tabulate a Title 13, Chapter 5 survey, census, or estimate;
7. Enhancing the data collected in a Title 13, Chapter 5 survey or census. For example:
 - Improving imputations for non-response;
 - Developing links across time or entities for data gathered in censuses and surveys authorized by Title 13, Chapter 5;
8. Identifying the limitations of, or improving, the underlying Business Register, Master Address File, and industrial and geographical classification schemes used to collect the data;
9. Identifying shortcomings of current data, collection programs and/or documenting new data collection needs;
10. Constructing, verifying, or improving the sampling frame for a census or survey authorized under Title 13, Chapter 5;

11. Preparing estimates of population and characteristics of population as authorized under Title 13, Chapter 5;
12. Developing a methodology for estimating non-response to a census or survey authorized under Title 13, Chapter 5;
13. Developing statistical weights for a survey authorized under Title 13, Chapter 5.

B. Scientific merit.

Projects without scientific merit will not contribute to existing knowledge. Evidence that a federal-funding agency, such as the National Science Foundation or the National Institutes of Health, using a peer review process, has approved a proposal for support meets this requirement. Proposals by graduate students to use Census Bureau data for a dissertation must include a statement by the student's advisor that the research methodology is appropriate for the proposed project and include such an advisor as a co-principal investigator.

C. Clear need for non-public data.

The proposal must demonstrate the need for and importance of access to non-public data. The proposal also should explain why publicly available data sources are not sufficient to meet the proposal's objectives.

D. Feasibility.

The proposal must show that the research is likely to be conducted successfully using the proposed methodology and requested data.

E. Risk of disclosure.

The researchers must accept all confidentiality protection and disclosure avoidance review requirements, including strict limits concerning how much and how often intermediate output can be taken out of the RDC. Output from all research projects must undergo and pass disclosure avoidance review. Therefore, the proposal must describe the probable outputs so that the Census Bureau can assess the likelihood that such outputs will pass disclosure avoidance review. Note that:

- Employees of the Census Bureau and sworn temporary staff, including those sworn users of Title 13, United States Code, confidential data at an RDC, are required by law to protect the confidentiality of the data collected under its authorizing legislation, 13 U.S.C. Sections 9, 23(c). Wrongful disclosure of confidential Title 13 data is punishable by fine not exceeding \$250,000.00 and/or imprisonment of no more than five years (cf. 13 U.S.C. Section 214; 18 U.S.C Section 3571).
- Tabular and graphical output presents a higher risk of disclosing confidential information than do coefficients from statistical models. Except in unusual circumstances, RDCs are reserved for projects that involve statistical or econometric modeling using economic and demographic microdata. RDCs are neither equipped nor designed to supplement the Census Bureau's existing data program operations by producing large-scale special tabulations from confidential Census Bureau microdata.
- Some data files are collected under the sponsorship of other agencies. In providing restricted access to these data, the Census Bureau must also adhere to all applicable laws and regulations under the sponsoring agency's authority. Researchers may be

required to sign non-disclosure documents of survey sponsors or other agencies that provide data for their research projects.

Proposal Review Process

Both Census Bureau and external experts on subject matter, datasets, and disclosure avoidance review all proposals. Proposals that would use data collected by the Census Bureau on behalf of another agency must be approved by that agency. In addition, all proposals undergo a review by the Census Bureau's Office of Analysis and Executive Support for compliance with the Census Bureau's policies. Proposals that clearly do not fall within existing Census Bureau policy guidelines will be referred to the Data Stewardship Executive Policy Committee for additional review. Relevant proposals must also abide by the requirements articulated in the *Administrative Records Handbook (DS-001)*, dated May 16, 2001.

The Census Bureau will inform the researchers about the outcome of the review process, and include a review synopsis, an explanation for the decision, and copies of the expert reviews.

Internal Census Bureau collaborative projects, that is, those that involve access to confidential Census Bureau data by both Census Bureau employees and Special Sworn Status employees, must meet the five review standards outlined above, whether they are carried out at an RDC or elsewhere at the Census Bureau. However, internal projects use a streamlined proposal review process. Internal projects must be submitted to the CES proposal management system in the same manner as external proposals. The division chief of the Census Bureau division sponsoring the project must attest in a memorandum to the Chief, CES that each proposed project meets the standards articulated above – benefit to Census Bureau programs, scientific merit, clear need for non-public data, feasibility, and acceptance of disclosure avoidance requirements. In addition, the sponsoring division must provide the following:

- Payment of RDC lab fees, if applicable;
- Any confidential data not yet available to the RDC network, along with associated documentation;
- Sufficient resources to store or process such data (if there are any unusual requirements of the proposed analysis); and
- Staff to assist the RDC researchers in their collaboration on this project, such as answering questions about the datasets, or assisting in disclosure avoidance review.

Special arrangements – subject to IT and physical security requirements – may be necessary should the Census Bureau collaborators wish to access newly created datasets, once on the RDC network, from the sponsoring division's location.

Reviewed proposals, external and internal, receive one of two ratings:

- **Approved.** The proposal successfully addresses all of the review standards mentioned above.

- **Not Approved.** The proposal fails to meet at least one review standard, and may be resubmitted as a new proposal only after suitable revision.

II. Data Linkage

Proposals that link records must comply with the requirements of applicable Census Bureau policies and procedures, primarily *Record Linkage Policy (DS-014)* dated February 5, 2004, and *Administrative Records Handbook (DS-001)* dated May 16, 2001. Projects proposing linkages may require approval of the Data Stewardship Executive Policy Committee (if that class of linkages has not yet been approved). If the linkages involve non-Title 13 data or Title 13 data collected for another agency, they require approval of the agency that provides or funds the collection of such data. Note that projects that require handling administrative records, linking, or similar services will be charged for the costs of such work.

Personal identifiers must remain under the control of those Census Bureau employees who are permitted access to them, under strictly controlled circumstances. For approved projects requiring linked person or household records, one of three procedures will be followed:

- (1) The linked files, created by Census Bureau employees with access to personal identifiers, will be stripped of identifiers and made available at the RDC;
- (2) Census Bureau employees with access to personal identifiers will replace them with “Protected Identification Keys” (PIKs). The files with PIKs will be made available to researchers so that the researchers can carry out the link; or
- (3) Researchers will receive some files with PIKs replacing personal identifiers. For other files, Census Bureau employees with access to personal identifiers may create crosswalks between PIKs and the unique (but not personal) identifiers on the files at the RDC so that the researcher can carry out the linkage.

III. Data Availability

CES does not have, and does not expect to have, all past and present Census Bureau confidential (Title 13) data available for RDC researchers. Researchers requesting data not currently available must consult with the relevant RDC Administrator about the procedures to make such data available, which may involve paying the costs of data and metadata development.

For projects using non-Title 13 confidential data that the researcher provides, only unmodified input files may be returned to the researcher; no modified input files or resulting matched files will be available outside the RDC environment. Researchers bringing data to an RDC for their project must provide documentation to CES from the data producer that the researchers are permitted to use the data; the documentation should also describe any restrictions on the use of the data by the researcher or others.

Public-use versions of any dataset may be included in a research proposal. It is the responsibility of the researcher to obtain such files and make them available to CES (unless they are already available). Proprietary private datasets can also be included in research

proposals provided the researchers provide evidence that the custodian of the data has approved their use by the researchers at an RDC and understands that linked versions cannot leave Census Bureau facilities. Researchers are not permitted simultaneous access to the confidential or linked versions and the public use version of the same dataset.

IV. Research Products and Disclosure Avoidance

The primary form of output from RDC research projects consists of coefficients from various types of regression or other statistical modeling procedures, plus certain statistics derived from these coefficients. With very few exceptions, such regression coefficients are considered safe for release because it is impossible to use the coefficients to reveal detailed information about individual respondents.

A second form of output consists of basic one- or two-dimensional tables of summary statistics (e.g., means, standard deviations). These tables are designed to give some basic information about the data that go into the analytical models; they are not the primary focus of the analysis. Most of these tabulations are safe because they are carried out at reasonably highly aggregated levels. Census Bureau employees perform primary and complementary disclosure avoidance analysis on these tabulations. If disclosure avoidance problems arise, the researcher is asked to collapse or suppress cells.

While disclosure avoidance review of regression statistics and the associated covariance matrixes and summary descriptive tables is relatively easy, it would be impractical for CES staff to review large, complex tabulations prior to removal from an RDC. Therefore, it is the Census Bureau's policy to limit the number and dimension of tabulations to be created at an RDC to those necessary to understand the statistical models' input data. If a researcher needs custom tabulations, that person must contact the sponsoring Census Bureau division directly, and expect to be charged an extra fee for complex disclosure avoidance review requirements. All custom tabulations must comply with the requirements of the Census Bureau's *Policy on Providing Custom Tabulations under Section 8(b) of Title 13 U.S.C. (DS-021)*, dated October 20, 2005.

All research output is subject to disclosure avoidance analysis. However, because research papers and other such output generated from outside-funded research projects are not considered official Census Bureau data products, the Census Bureau does not impose its publication standards on researchers' output. Researchers are required to put a disclaimer to this effect on any papers or other publicly distributed output. The required disclaimer is worded as follows:

This research was carried out at a U.S. Census Bureau Research Data Center facility [the location may be specified]. The results and conclusions of the paper are those of the author(s) and do not indicate concurrence by the Census Bureau. These results have been screened to avoid revealing confidential data.

All research papers using data available only at an RDC must be submitted for inclusion in the Center for Economic Studies Discussion Paper series.